Article Watch: July 2019

Clive A. Slaughter

This column highlights recently published articles that are of interest to the readership of this publication. We encourage ABRF members to forward information on articles they feel are important and useful to Clive Slaughter, MCG-UGA Medical Partnership, 1425 Prince Avenue, Athens, GA 30606, USA. Tel: (706) 713-2216; Fax: (706) 713-2221; E-mail: cslaught@uga.edu, or to any member of the editorial board. Article summaries reflect the reviewer's opinions and not necessarily those of the Association.

NUCLEIC ACID SEQUENCING AND GENOTYPING

Liu Y, Siejka-Zielińska P, Velikova G, Bi Y, Yuan F, Tomkova M, Bai C, Chen L, Schuster-Böckler B, Song C-X. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nature Biotechnology* 37;2019: 424-429.

Bisulfite sequencing, the gold standard of methods for base-level detection and quantification of 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), works by treating DNA with bisulfite to convert unmethylated cytosine to uracil while leaving 5mC and/or 5hmC unchanged. Because uracil is read as thymine during PCR amplification, the modification of cytosine can be inferred on the basis of the resulting C-to-T transition. However, the method has drawbacks. The chemically harsh bisulfite treatment degrades DNA by depyrimidination, which limits its utility in sample-limited studies. Furthermore, the complete conversion of cytosine to thymine severely reduces sequence complexity, which results in poor sequencing quality. Liu *et al.* now describe a new bisulfite-free methodology to avoid these problems by using mild conditions to directly detect 5mC and 5hmC at base-level resolution. The authors first convert 5mC and 5hmC to 5-carboxycytosine (5caC) with a ten-eleven translocation (TET) dioxygenase enzyme, then reduce 5caC to dihydrouracil (DHU) with pyridine borane, or its derivative 2-picoline borane, in a previously unknown decarboxylation/deamidation reaction. Both reactions are mild and non-destructive. DHU is recognized as thymine by DNA and RNA polymerases. 5mC and 5hmC are therefore detected as a C-to-T transition during sequencing. The authors name this methodology TET-assisted pyridine-borane sequencing (TAPS). Notably, 5hmC can be protected from TET oxidation and borane reduction by addition of glucose using β-glucosyltransferase. This intervention yields a process for localizing 5hmC that the authors call TAPSβ. TAPS detects 5mC and 5hmC with high sensitivity and specificity. It outperforms bisulfite sequencing by providing higher, more even coverage of the methylome with less starting material at half the sequencing cost. Replacement of bisulfite sequencing by TAPS as the gold standard for methylome analysis is anticipated. The new method is also potentially compatible with PCR-free DNA sequencing methods and with long-read sequencing technologies.

Haile S, Corbett R D, Bilobram S, Bye M H, Kirk H, Pandoh P, Trinh E, MacLeod T, McDonald H, Bala M, Miller D, Novik K, Coope R J, Moore R A, Zhao Y, Mungall A J, Ma Y, Holt R A, Jones S J, Marra M A. Sources of erroneous sequences and artifact chimeric reads in next-generation sequencing of genomic DNA from formalin-fixed paraffin-embedded samples. *Nucleic acids research* 47; 2019:e12-e12.

Haile *et al.* investigate one class of artifacts, namely, chimeric reads, that are encountered at elevated levels in DNA sequences derived from formalin-fixed, paraffin-embedded tissue samples stored in pathology laboratories. Such samples constitute a potentially valuable resource in the study of the mechanisms of pathogenesis. By comparing 38 matched fresh-frozen and formalin-fixed, paraffin-embedded cancer samples, the authors collect evidence that many of the chimeric artifacts are derived from non-contiguous single-stranded regions of denatured DNA that become linked together during end-repair as a result of short stretches of sequence complementarity. The authors show that removal of single-stranded DNA fragments by treatment with S1 nuclease reduces the error rates. With this treatment, they also observe reduced GC-bias, more uniform sequence coverage and fewer false-positive single nucleotide variants and copy number variants.

doi: 10.7171/jbt.19-3002-002

MACROMOLECULAR SYNTHESIS AND SYNTHETIC BIOLOGY

Hoshika S, Leal N A, Kim M-J, Kim M-S, Karalkar N B, Kim H-J, Bates A M, Watkins N E, SantaLucia H A, Meyer A J, DasGupta S, Piccirilli J A, Ellington A D, SantaLucia J, Georgiadis M M, Benner S A. Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* 363;2019:884-887.

Hoshika *et al.* have designed 2 nonphysiologic base pairs to incorporate into DNA, creating a polymer that functions with an 8-letter genetic code. In the standard base pairs, guanine:cytosine (G:C) and adenine:thymine (A:T), a purine is paired with a pyrimidine. In the new base pairs, a purine analog is paired with a pyrimidine analog. The purine analogs are labeled "P" and "B." The pyrimidine analogs are labeled "Z" and "S." These bases are sterically designed to produce the pairings P:Z and B:S. Unlike previous studies in which nonstandard bases are incorporated into DNA, the new base pair interactions are stabilized by hydrogen bonding. The authors term the polymer incorporating the novel bases "hachimoji" DNA (*hachi* = 8, *moji* = letters). Using solid-phase phosphoramidite chemistry, they synthesize 94 polymers incorporating A, T, G, C, P, B, Z and S and measure thermodynamic parameters that are shown accurately to predict duplex melting temperatures. This indicates that hachimoji DNA reproduces the molecular recognition behavior of standard DNA and thus acts as an informational system. The authors further determined high-resolution crystal structures for 3 (16-mer) hachimoji duplexes and showed that these duplexes adopt double helical structures of the B-form with small differences in pairing characteristics well within the range for natural DNA pairings. Hachimoji DNA therefore meets the Schrödinger requirement for the bases to fit into an aperiodic crystal regardless of sequence, indicating that the structure will support Darwinian evolution. The authors then developed a thermostable T7 RNA polymerase variant capable of transcribing hachimoji DNA into hachimoji RNA. They designed a hachimoji variant of the spinach fluorescent RNA aptamer. When folded, the RNA aptamer binds 3,5-difluoro-4-hydroxybenzylidene imidazolinone, a molecule that fluoresces green when bound. They showed that hachimoji DNA can be transcribed into a hashimoji RNA that fluoresces green. In all respects thus far tested, hashimoji nucleic acids therefore meet the criteria for molecular structures capable of supporting life.

Reinkemeier C D, Girona G E, Lemke E A. Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. *Science* 363;2019:eaaw2644.

The work described in this paper sets the goal of engineering a eukaryotic cell that incorporates both the endogenous translation system and a parallel translation system with fully orthologous components, such that the cell synthesizes one protein that incorporates the non-canonical amino acid but synthesizes all the other proteins normally. Although genetic code expansion has been described previously, for example by supplying an orthogonal tRNA/tRNA synthase pair to reprogram a stop codon, the noncanonical amino acid encoded by the reprogrammed codon is incorporated whenever that codon is encountered. The present work sets the specificity bar much higher by demanding that only one protein incorporates the non-canonical amino acid. The authors approach their task by sequestering the components of the novel translation system into a membraneless orthogonal translation organelle. They bring the modified tRNA/tRNA synthase and the translated mRNA of choice into close proximity by a combination of phase separation and spatial targeting. (For another application of phase separation, see the discussion of coacervates in the summary of the paper by Koolivand *et al.* below.) The authors engineer the mRNA with sequences that bind to the RNA-binding domain of a phage capsid protein. The tRNA synthase (which in turn binds it to cognate tRNA) is fused to this capsid protein. The capsid protein and the tRNA synthase are also fused to long intrinsically disordered domains found on proteins that generate highly concentrated assemblies in cells. These disordered sequences aggregate to form a dense phase/condensate constituting a membraneless organelle. Cells containing these species efficiently and selectively decode mRNAs for the protein chosen for incorporation of the noncanonical amino acid. The results indicate that the methodology will provide a general platform for generating semisynthetic eukaryotic cells and organisms.

MASS SPECTROMETRY

Sinclair I, Bachman M, Addison D, Rohman M, Murray D C, Davies G, Mouchet E, Tonge M E, Stearns R G, Ghislain L, Datwani S S, Majlof L, Hall E, Jones G R, Hoyes E, Olechno J, Ellson R N, Barran P E, Pringle S D, Morris M R, Wingfield J. Acoustic mist ionization platform for direct and contactless ultrahigh-throughput mass spectrometry analysis of liquid samples. *Analytical Chemistry* 91;2019:3790-3794.

Sinclair *et al.* present instrumentation and operational improvements to a previously described inlet system for electrospray mass spectrometry that is capable of very high rates of sample introduction: up to 3 individual samples per

second. The principle of operation, referred to as acoustic mist ionization (AMI), relies upon the contactless delivery of acoustic energy to transfer samples from a microtiter plate directly into the mass spectrometer. An external transducer emits pulses of acoustic energy into a sample in a 364-well plate located on a moving *XY*-stage. High voltage is applied to a charging cone suspended directly above the sample plate. The charging cone induces charge separation within the sample volume, and a mound that forms on the sample liquid meniscus adjacent to the cone ejects charged droplets (5.5 μm diameter, 87 fL volume) that spray through the cone into a heated transfer tube leading into the mass spectrometer ion source. High voltage and electrically insulated wells are required for AMI to work, thus distinguishing it from solvent-assisted inlet ionization (SAII) and surface acoustic wave nebulization (SAWN) methods. The authors optimize the conditions for acoustic mist generation in the present publication and describe hardware changes to improve efficiency. They achieve sample throughput of 100,000 samples per day on a single mass spectrometer with plate-handling automation, a rate suitable for application in drug discovery. In a test application of the technology, 2.75 million wells (7164 plates) were screened in only 7 wk with a single mass spectrometer.

Kafader J O, Melani R D, Senko M W, Makarov A A, Kelleher N L, Compton P D. Measurement of individual ions sharply increases the resolution of Orbitrap mass spectra of proteins. *Analytical Chemistry* 91;2019:2776-2783.

In Fourier transform–based mass spectrometers such as Orbitraps, resolution is ultimately limited by the decay rate of the time-domain signal. Signal decay results from analyte collisions with background gas, which may cause dampening of ion motion or ion dissociation. These effects worsen with increasing analyte mass, which is associated with an increase in collisional cross-section. The very large number of charges on protein ions may also exacerbate space-charging, which leads to field distortion. To circumvent these problems, the present paper describes methodology based on principles of single ion detection and applies them to protein mass spectrometry in an Orbitrap mass analyzer. Orbitraps have been proven capable of detecting single ion events for proteins as small as myoglobin (charge +20). Only a few ions are collected per scan. Spectra are acquired in profile mode and filtered to exclude noise and decayed ions. Centroids are then determined, and the centroid data are summed to produce a histogram of ion distribution on the m/z scale. The authors apply this procedure to proteins with molecular weight from 8 to 150 kDa. They achieve a resolving power of 677,000 for transients of carbonic anhydrase (29 kDa) with a duration of only ~250 ms. Such isotopic resolution allows robust assignment of overlapping distributions, including adducts and degradation products.

FUNCTIONAL GENOMICS AND PROTEOMICS

Wienert B, Wyman S K, Richardson C D, Yeh C D, Akcakaya P, Porritt M J, Morlock M, Vu J T, Kazane K R, Watry H L, Judge L M, Conklin B R, Maresca M, Corn J E. Unbiased detection of CRISPR off-targets *in vivo* using DISCOVER-Seq. *Science* 364; 2019:286-289.

Clustered regularly interspaced short palindromic repeat (CRISPR)-based gene editing has provided an incomparably powerful set of tools for the directed introduction of sequence changes in germline DNA. Both the interpretation of experimental results and the success of putative therapeutic applications depend on the specificity for the targeted gene with which sequence changes are introduced. Consequently, much effort is presently directed toward the design of optimal experimental methods to assess target specificity. Wienert *et al.* have developed a new method by which to assess the target specificity with which active Cas endonucleases introduce double-stranded breaks in DNA under targeting by a guide RNA for the purpose of gene inactivation. Upon introduction of a double-stranded break, cells recruit endogenous repair enzymes to the damaged site. The authors make use of this process by identifying the location at which such recruitment takes place. They localize the DNA sites at which the MRE11-RAD50-NBS1 repair complex (the so-called MRN complex) binds. This localization is achieved by CHIP-Seq analysis based on the MRE11 component. MRE11 is broadly expressed, and an antibody that reacts with both the mouse and human protein is available.

Jin S, Zong Y, Gao Q, Zhu Z, Wang Y, Qin P, Liang C, Wang D, Qiu J-L, Zhang F, Gao C. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* 364;2019:292-295.

Zuo E, Sun Y, Wei W, Yuan T, Ying W, Sun H, Yuan L, Steinmetz L M, Li Y, Yang H. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* 364;2019:289-292.

Two groups newly address the specificity of CRISPR base editing, a process in which a guide RNA targets a catalytically inactive Cas enzyme conjugated to a nucleotide

deaminase for the purpose of converting one nucleotide to another. Off-target activity is especially difficult to establish in this context because it occurs against a background of naturally occurring sequence variability. Zuo *et al.* determine off-target activity by injecting a Cas base editor and guide RNA into one cell of a 2-cell embryo, along with a fluorescent maker to indicate which is the treated cell and its progeny. They then sort treated from untreated progeny cells on the basis of fluorescence and perform whole-genome sequencing to detect sequence differences. In the absence of editing, the sequences should be identical. Jin *et al.* use a similar approach in rice plants. They introduce editing components into single cells and then sequencing the rice plants produced therefrom. Both groups find a substantially higher frequency of off-target editing with the cytosine base editor than with the adenosine base editor. The difference has been ascribed to the properties of the two deaminases employed (Kempton HR, Qi LS. *Science* 364; 2019:234-236). The cytosine base editor uses a cytosine deaminase that can bind to single-stranded DNA independently of Cas9, whereas the adenosine deaminase lacks the ability to bind on its own. These results indicate a need for further developmental work to improve the specificity of cytosine editing.

PROTEOMICS

The M, Käll L. Integrated identification and quantification error probabilities for shotgun proteomics. *Molecular & Cellular Proteomics* 18;2019: 561-570.

Typical shotgun proteomic pipelines for identifying differentially expressed proteins attempt to control error rates by imposing false discovery thresholds or other heuristic cutoff values at each stage of the discovery pipeline, for example, requiring a certain number of peptide spectrum matches for acceptable protein identification, imposing a 5% false-discovery rate on protein identifications, and requiring a certain number of peptides as evidence of differential protein expression. When a list is compiled of candidate proteins that satisfy a threshold at one step in the pipeline, the probability that it contains false-positives is often ignored in subsequent steps. However, the goal should be to estimate the combined probability that a protein is correctly identified, correctly quantified, and correctly labeled as differentially expressed. Errors introduced by poor strategies for imputation of missing values or by ignoring multiple hypothesis testing may also cause severe increases in false discovery rates that often go unrecognized. The present paper reports a probabilistic graphical method based on Bayesian statistics for propagating error information along the entire identification pipeline. Specifically, it replaces statistically unsound foldchange cutoffs by posterior probabilities that combine both identification error rates and differential expression error rates. The authors show through application to engineered and clinical datasets that this approach yields not only improved control of false discovery rates but also improved sensitivity of discovery of significant proteins. The model executes in minutes and is made freely available *via* the internet.

Koolivand A, Azizi M, O'Brien A, Khaledi M G. Coacervation of lipid bilayer in natural cell membranes for extraction, fractionation, and enrichment of proteins in proteomics studies. *Journal of Proteome Research* 18;2019:1595-1606.

Coacervates are dense liquid droplets composed of amphipathic molecules that separate from the dilute phase through liquid-liquid phase separation. This process is driven by interactions between the amphiphiles. Koolivand *et al.* induce coacervation of the lipid components of yeast cell membranes by adding a low concentration of a water-miscible fluorinated alcohol—in the present work, 16% hexafluoroisopropanol (HFIP) in water. HFIP is added to a tube containing a yeast cell lysate in 50 mM aqueous ammonium bicarbonate or 6.5 M urea, and the resulting dense coacervate separates from the aqueous layer as a distinct phase at the bottom of the tube following brief, low-speed centrifugation. The coacervate is shown to have a total concentration of endogenous lipids of 15 mg/ml under the conditions employed. Because the coacervate is consequently hydrophobic in nature, it solubilizes and extracts hydrophobic proteins. Upon removal of the aqueous layer, HFIP may be removed from the coacervate layer by evaporation under a stream of nitrogen, and the proteins it contains may then be prepared for analysis by LC-MS/MS. Integral membrane proteins are shown to separate into the coacervate layer, whereas anchored membrane proteins and cell wall proteins stay in the aqueous layer. This method for preparing integral membrane proteins is easy and free of detergent. It is favorable for the identification of low-abundance proteins and integral membrane proteins of organelles.

PROTEIN CHARACTERIZATION

Egloff P, Zimmermann I, Arnold F M, Hutter C A J, Morger D, Opitz L, Poveda L, Keserue H-A, Panse C, Roschitzki B, Seeger M A. Engineered peptide barcodes for in-depth analyses of binding protein libraries. *Nature Methods* 16;2019:421-428.

Egloff *et al.* describe a method for selection of proteins that bind a particular ligand from a library containing an

ensemble of thousands of proteins at once, without the need to handle individual proteins at any stage of the procedure. The method uses genetically encoded peptide barcodes to identify the different proteins in the library. The barcodes are designed for optimal detection and identification by liquid chromatography–tandem mass spectrometry (LC-MS/MS). They are 11–15 residues in length and contain 2 randomized regions that produce a theoretical library diversity of 5.3×10^8 members. A library of incipient binders is cloned into a plasmid that harbors the barcode library. The number of alternative barcodes per binder is >30, and barcodes identify binders unambiguously. The nested library is subjected to high-throughput DNA sequencing to assign barcodes to binders. After selection of proteins that bind to the ligand, the binding proteins are digested to release their barcode peptides, which are detected preferentially by LC-MS/MS against the background of peptides from the binding proteins themselves because of the favorable sequence design of the barcodes. Utility of the method is illustrated in 3 applications: ranking of nanobodies according to their binding to a protein antigen; sampling the diversity of camelid nanobodies against a protein antigen directly from B cells of immunized alpacas; and directly detecting rare binders in the immune repertoire of nanobodies against an integral membrane protein of *Legionella* in a cellular context.

Touti F, Gates Z P, Bandyopdhyay A, Lautrette G, Pentelute B L. In-solution enrichment identifies peptide inhibitors of protein-protein interactions. *Nature Chemical Biology* 15;2019:410-418.

To address a problem representing the converse of that studied by Egloff *et al.* above, Touti *et al.* describe methodology for improving the affinity for a protein target of peptides that may be employed as pharmacologic inhibitors of endogenous interactions. The authors develop a method for affinity-based selection of peptides binding to a protein target from a library of 10^3–10^6 synthetic peptides that also incorporate noncanonical amino acids. The library randomizes select residues within the sequence of a known peptide binder using a split-and-pool synthesis technique. After binding to the protein target, the method uses high-performance size-exclusion chromatography to select the peptides that bind to the target protein and then identifies these peptides by LC-MS/MS sequencing. No reporter labels or encoding tags are required. The authors test their method in 2 applications. They realize gains in affinity of 100× for inhibitors of MDM2-TP53 interaction and 30× for inhibitors of dimerization of HIV capsid protein C-terminal domain. The identified high-affinity binders are then cyclized to render them cell-permeable for biologic assay. This approach is expected to contribute to the streamlining of drug discovery pipelines.

CELL BIOLOGY AND TISSUE ENGINEERING

Liu Y, Mi Y, Mueller T, Kreibich S, Williams E G, Van Drogen A, Borel C, Frank M, Germain P-L, Bludau I, Mehnert M, Seifert M, Emmenlauer M, Sorg I, Bezrukov F, Bena F S, Zhou H, Dehio C, Testa G, Saez-Rodriguez J, Antonarakis S E, Hardt W-D, Aebersold R. Multiomic measurements of heterogeneity in HeLa cells across laboratories. *Nature Biotechnology* 37;2019:314-322.

More than 100,000 papers have been published on studies that have used, or referred to, HeLa cells. The genomic instability of HeLa is widely known, but the phenotypic impact of variability among HeLa cell lines and the impact of differences that accumulate as the cells are passaged have not been systematically investigated. Liu *et al.* report the degree of molecular and phenotypic variability among HeLa cell line variants obtained from 13 international laboratories. They determine genomic copy number profiles, use sequential-window acquisition of all theoretical fragments (SWATH) mass spectrometry (SWATH-MS) to quantify proteomic profiles of 5000 proteins (about half of the proteins expressed in HeLa cells), measure protein loss rates as a proxy for protein turnover using stable isotope labeling with amino acids in cell culture (SILAC)-SWATH-MS, and document phenotypic characteristics that include cell doubling time and response to *Salmonella*. The cells from different sources are passaged under identical conditions to control for technical variations. The results reveal substantial variability between HeLa variants. Notably, CCL2, the "original" HeLa cell line, and Kyoto varieties of HeLa are consistently different at every level, including the functional level. They are as distinct as cancer cell lines from different tissue types. Furthermore, CCL2 cells diverge during 3 mo in continuous culture under uniform conditions in the same laboratory. They accumulate substantial changes in copy-number variants (CNVs) and differential expression of ~6% of genes. The proteomic diversity can be attributed in part to CNV changes. Most of the proteomic variance is concordant with variation in mRNA levels, but turnover plays a significant role for some proteins. The authors suggest measures to minimize the effect of all this variability on biologic research results. They advocate documentation of the origin of HeLa cells used, annotation of the passages of cells during a study, and repetition of important results with different samples of the cell line. This study has important implications for the interpretation and reproducibility of

research with cultured cell lines generally. We may therefore anticipate increasing use of transcriptomic and possibly proteomic profiling to document the status of cells used in biologic research.

Skinnider M A, Squair J W, Foster L J. Evaluating measures of association for single-cell transcriptomics. *Nature Methods* 16;2019:381-386.

Single-cell RNA sequencing is an incomparably powerful tool for analysis of differential gene expression, function of signaling pathways, identification of cell types, and ascertainment of developmental relationships. However, the statistical analysis of single-cell transcriptomic data presents significant challenges. Many measures of association have been proposed, but none has emerged as the gold standard. Skinnider *et al.* compare the performance of 17 measures of association for tasks that include gene network analysis, cell clustering, and identification of disease genes. They evaluate performance on the basis of the functional coherence of the constructed gene networks, the reproducibility of the constructed networks across datasets, and the correspondence of these networks with networks based on other criteria, such as protein-protein interaction, cellular signaling, and metabolic relationships. The authors also assess the reviewed methods for their ability to assign cell types and disease associations. Measures of goodness-of-fit to proportionality, ρ_p and ϕ_s, respectively, are shown to perform better across the diverse tasks assessed than measures of association (including Pearson correlation, Euclidian distance, mutual information, and cosine distance). Nevertheless, the authors comment that predictive power may be enhanced significantly by integrating the single-cell RNA sequencing data with other data types, such as proteomic and epigenomic data. Investigators may further wish to consider adopting measures of proportionality in place of conventional measures of association for additional tasks in which biologic information is extracted from "omic" datasets.

Fischer D S, Fiedler A K, Kernfeld E M, Genga R M J, Bastidas-Ponce A, Bakhti M, Lickert H, Hasenauer J, Maehr R, Theis F J. Inferring population dynamics from single-cell RNA-sequencing time series data. *Nature Biotechnology* 37;2019:461-468.

Single-cell RNA sequencing provides essentially static "snapshots" of cell states. Developmental trajectories—the pathways by which cells transition from one state to another—may be inferred from these snapshots by acquiring data at multiple time points. The present paper considers the questions of how fast and in what numbers cells move along their trajectories. These questions of population dynamics are addressed by combining single-cell snapshot data with estimates of population size. The proposed mathematical framework allows changes in cell proliferation rate, death rate, and differentiation to be disentangled. Application of the methodology is illustrated in studies of T-cell maturation and pancreatic β-cell maturation.

IMAGING

Ghosh R P, Franklin J M, Draper W E, Shi Q, Beltran B, Spakowitz A J, Liphardt J T. A fluorogenic array for temporally unlimited single-molecule tracking. *Nature Chemical Biology* 15;2019:401-409.

Ghosh *et al.* describe new reagents for the tracking of a single molecule inside cells for periods ranging from milliseconds to hours. The authors employ the 26-fold fluorescence enhancement of green fluorescent protein (GFP) that occurs when it interacts with a binding protein, which in this case is a camelid nanobody. Cells synthesize the molecule of interest tagged with a concatenated array of these nanobodies to which up to 24 GFP molecules can be bound. The authors track cytoplasmic motors, such as kinesin, and integrins for thousands of frames. To study nuclear proteins, they designed a version of the array tag with repeated copies of a nuclear import signal. They use it to track single histones at 0.5 Hz for more than 1 h. They also develop a tag based on an array of anti-dihydrofolate reductase nanobodies that can bind cognate dihydrofolate reductase conjugated to a fluorophore (in this case mCherry, although many alternatives could be used). This construct is employed with the GFP array binder for simultaneous dual-color imaging. These tags have large molecular size, but integrin β1 and histone H2B are found to be tolerant of the payload in functional assays. The constructs are expected to help address questions about single molecules related to assembly, targeting, motility, turnover, and cell division.

Gao R, Asano S M, Upadhyayula S, Pisarev I, Milkie D E, Liu T-L, Singh V, Graves A, Huynh G H, Zhao Y, Bogovic J, Colonell J, Ott C M, Zugates C, Tappan S, Rodriguez A, Mosaliganti K R, Sheu S-H, Pasolli H A, Pang S, Xu C S, Megason S G, Hess H, Lippincott-Schwartz J, Hantman A, Rubin G M, Kirchhausen T, Saalfeld S, Aso Y, Boyden E S, Betzig E. Cortical column and whole-brain imaging with molecular contrast and nanoscale resolution. *Science* 363;2019:eaau8302.

Gao *et al.* illustrate the capabilities of methodology that combines expansion microscopy, lattice light-sheet microscopy, and computational methods for image processing on the terabyte scale to deepen knowledge of the functional architecture of nervous systems. The investigators chemically link fluorophore-conjugated antibodies, fluorescent proteins, or both, to a polyacrylamide/polyacrylate gel formed of components infused into the tissue. After unfixed proteins and lipids are removed to render the tissue transparent, it is isotropically swollen with water to 4 times its original size while retaining the relative distribution of fluorescent tags. The expanded sample is imaged using lattice light-sheet microscopy, a technique that rapidly scans with a thin sheet of light to minimize photobleaching and out-of-focus background. This provides ~60 × 60 × 90-nm resolution at 4× expansion. An entire *Drosophila* brain (340 × 660 × 90 μm) can be imaged with single-molecule sensitivity at acquisition rates estimated to be ~700 times faster than existing high-speed superresolution fluorescence microscopy, in ~63 h. The process generates multiterabyte datasets. The authors develop a computing pipeline that corrects for intensity variations across the light-sheet and slight swelling or shrinking of the sample during acquisition. Using this methodology, the authors investigate the subcellular spatial relationships of neuronal proteins on the scale of nanometers across the entire thickness of the mouse cerebral cortex or the entire volume of the *Drosophila* brain on the scale of hundreds of micrometers. The structures they investigate include synaptic proteins on dendritic spines, myelination of axons, and presynaptic protein densities on dopaminergic neurons. The methodology therefore promises to contribute to the integration of the high-resolution data necessary to understand neural behavior with the long-range mapping necessary to understand neural circuits.